

SEQUENCE LISTING

<110> Diversa Corporation
 Morgan, Brian
 Burk, Mark
 Levin, Michael
 Zhu, Zoulin
 Chaplin, Jennifer
 Kustedjo, Karen
 Huang, Zilin
 Greenberg, William

<120> METHODS FOR MAKING SIMVASTATIN AND INTERMEDIATES

<130> 564462012840

<140> Not Yet Assigned.

<141> 2004-10-20

<150> US 60/542,100

<151> 2004-02-04

<150> US 60/513,237

<151> 2003-10-21

<160> 6

<170> PatentIn version 3.1

<210> 1

<211> 1629

<212> DNA

<213> Unknown

<220>

<223> Environmental

<400> 1
 atgagccttt gcgtcattcg attcatcgcc ggaactttgg tacttgtggc gtcagtggaa 60
 tcggcagttg ctcaacaagc gtgtgctgac ctgatgggcc tcgagctgcc gtataacaacg 120
 ataacgtccg ctgcagtggc taccgagggc ccaatccac agccggcgat ctttggaagc 180
 actgacccca ttgtggctcc agagcgatgt gaagtgcggg cggtcacgcg ccctacgaag 240
 gactccgaga ttcgaatcga gctctggctg ccgctctccg gatggaacgg aaagtatcta 300
 caaattggtg gcggtggctg ggctgggttc atcaatcgaa cggggctgat aggccctctt 360
 cagcgcgggt atgccgtagc cgcaaccgac aatggccata tcagcgaagg tttggtgcct 420
 gacgcctcct gggctatcgg ccatccgcaa aagctgatcg atttcggtta tcgcgcctg 480
 cacgaaacaa gtgttcaggc caaagctatc ctgcgcgcct actttggccg cggtcaggat 540
 ctgagctact tcagcggttg ttctaattgg ggacgcgagg ctctcatgga ggcgacgcgc 600
 tatccggaag atttcgaagg catcatcgcc ggagcgcccc cgaacaattg gtcgcgcctg 660

```

tttacggggt ttgtgtggaa tgaacgcgcg ttggcggacg atccaattcc tcctgccaag      720
ttgacagcga ttcaggcggc ggcaattgct gcgtgtgata cgctggacgg tgttgaggac      780
gggctcatcg aaaaccacg agcgtgtagc ttcgatccgc gttcaatggc ctgtacagcc      840
gatgatgcct ctgactgtct gacagaagga caggctcgca cgctacacag gatatatagc      900
ggcccaacca atcctcggac cggtgagcga atctttccag gctatccgat gggcaccgaa      960
gccgtgccgg gcggatgggt accgtggatc gtgtccgcga gctccgaagt tccgagcata     1020
caagcaagct ttggcaactc ctattacggg cacgcggtct tcgagcaatc gaactgggat     1080
ttcaggacgt tggatttcga ccaggacgtt gcgtttggcg atgcgaaggc ggggccgggtg     1140
ctcaatgcca cgaaccccgga tctgcgttcg tttcgcgcga atggcggcaa actgattcag     1200
tatcatggct ggggcgatgc agccattacg gcttttagtt cgatcgacta ctacgagaac     1260
gtgcgcgcct tcctcgatcg cttccccgac ccccggaagcg agaacacgga tatcgacggt     1320
ttctatcgcc tgttcttggt tccgggcatg ggacattgct ccggcgggat cggcccaagt     1380
agctttggca atggcttccg ttccgcacgt acggatgccg agcacgacct actctccgcc     1440
cttgaggcat ggggtggagcg agacacggcc ccggagagat tgatcggaac ggggacggcc     1500
gtaggcgacc caaccgcgac tctgacgcgt ccgctatgcc catatccgcg gacggcacgg     1560
tatctcgaa gcggaactc aaatgatgcg gccaaacttcg agtgcgccct gcccgctggc     1620
gtgcagtag

```

<210> 2
 <211> 542
 <212> PRT
 <213> Unkown

<220>
 <223> Environmental

<220>
 <221> SIGNAL
 <222> (1)...(24)

<400> 2
 Met Ser Leu Cys Val Ile Arg Phe Ile Ala Gly Thr Leu Val Leu Val
 1 5 10 15

Ala Ser Val Glu Ser Ala Val Ala Gln Gln Ala Cys Ala Asp Leu Met
 20 25 30

Gly Leu Glu Leu Pro Tyr Thr Thr Ile Thr Ser Ala Ala Val Ala Thr
 35 40 45

Glu Gly Pro Ile Pro Gln Pro Ala Ile Phe Gly Ser Thr Asp Pro Ile
50 55 60

Val Ala Pro Glu Arg Cys Glu Val Arg Ala Val Thr Arg Pro Thr Lys
65 70 75 80

Asp Ser Glu Ile Arg Ile Glu Leu Trp Leu Pro Leu Ser Gly Trp Asn
85 90 95

Gly Lys Tyr Leu Gln Ile Gly Ser Gly Gly Trp Ala Gly Ser Ile Asn
100 105 110

Arg Thr Gly Leu Ile Gly Pro Leu Gln Arg Gly Tyr Ala Val Ala Ala
115 120 125

Thr Asp Asn Gly His Ile Ser Glu Gly Leu Val Pro Asp Ala Ser Trp
130 135 140

Ala Ile Gly His Pro Gln Lys Leu Ile Asp Phe Gly Tyr Arg Ala Val
145 150 155 160

His Glu Thr Ser Val Gln Ala Lys Ala Ile Leu Arg Ala Tyr Phe Gly
165 170 175

Arg Gly Gln Asp Leu Ser Tyr Phe Ser Gly Cys Ser Asn Gly Gly Arg
180 185 190

Glu Ala Leu Met Glu Ala Gln Arg Tyr Pro Glu Asp Phe Glu Gly Ile
195 200 205

Ile Ala Gly Ala Pro Ala Asn Asn Trp Ser Arg Leu Phe Thr Gly Phe
210 215 220

Val Trp Asn Glu Arg Ala Leu Ala Asp Asp Pro Ile Pro Pro Ala Lys
225 230 235 240

Leu Thr Ala Ile Gln Ala Ala Ala Ile Ala Ala Cys Asp Thr Leu Asp
245 250 255

Gly Val Glu Asp Gly Leu Ile Glu Asn Pro Arg Ala Cys Ser Phe Asp
260 265 270

Pro Arg Ser Met Val Cys Thr Ala Asp Asp Ala Ser Asp Cys Leu Thr

275	280	285
Glu Gly Gln Val Ala Thr Leu His Arg Ile Tyr Ser Gly Pro Thr Asn		
290	295	300
Pro Arg Thr Gly Glu Arg Ile Phe Pro Gly Tyr Pro Met Gly Thr Glu		
305	310	315
Ala Val Pro Gly Gly Trp Val Pro Trp Ile Val Ser Ala Ser Ser Glu		
325	330	335
Val Pro Ser Ile Gln Ala Ser Phe Gly Asn Ser Tyr Tyr Gly His Ala		
340	345	350
Val Phe Glu Gln Ser Asn Trp Asp Phe Arg Thr Leu Asp Phe Asp Gln		
355	360	365
Asp Val Ala Phe Gly Asp Ala Lys Ala Gly Pro Val Leu Asn Ala Thr		
370	375	380
Asn Pro Asp Leu Arg Ser Phe Arg Ala Asn Gly Gly Lys Leu Ile Gln		
385	390	395
Tyr His Gly Trp Gly Asp Ala Ala Ile Thr Ala Phe Ser Ser Ile Asp		
405	410	415
Tyr Tyr Glu Asn Val Arg Ala Phe Leu Asp Arg Phe Pro Asp Pro Arg		
420	425	430
Ser Glu Asn Thr Asp Ile Asp Gly Phe Tyr Arg Leu Phe Leu Val Pro		
435	440	445
Gly Met Gly His Cys Ser Gly Gly Ile Gly Pro Ser Ser Phe Gly Asn		
450	455	460
Gly Phe Arg Ser Ala Arg Thr Asp Ala Glu His Asp Leu Leu Ser Ala		
465	470	475
Leu Glu Ala Trp Val Glu Arg Asp Thr Ala Pro Glu Arg Leu Ile Gly		
485	490	495
Thr Gly Thr Ala Val Gly Asp Pro Thr Ala Thr Leu Thr Arg Pro Leu		
500	505	510

Cys Pro Tyr Pro Arg Thr Ala Arg Tyr Leu Gly Ser Gly Asn Ser Asn
 515 520 525

Asp Ala Ala Asn Phe Glu Cys Ala Leu Pro Ala Gly Val Gln
 530 535 540

<210> 3
 <211> 1209
 <212> DNA
 <213> Unknown

<220>
 <223> Environmental

<400> 3
 atggaaatcc atggtacatg cgacccaaag tttcacttgg tgcggcagga gtttgaacga 60
 aatttgcgtg agcgcggcga agtaggagcg tccgtttgcg tcacgttgca cggcgaaacc 120
 gtagtggact tgtggggcgg catggcgcgt gccgacactc agacgccatg gacggcggag 180
 acggtcagta ttgttttttc ctccacaaa ggcgcaacgg cactctgcgc ccatatgctg 240
 gcgtcacgcg gccaaactgga tcttgatgca ccagtcgcca cctactggcc ggaatttgcc 300
 caagccggca aagctcgcat cccggtgaaa atgctcttga accatcaagc tggctctccct 360
 gccgtacgga caccgctgcc ccagggtgcc tacgctgact gggaactgat ggtcaatacg 420
 ttggccaagg aagagccgtt ttgggaacct ggcacccgca acggctatca tgcgctcacc 480
 atgggggtggc tgggtgggaga agtgggtgca cgtgtctctg gtaagtcgct tgggacattc 540
 ttccaagagg agatcgccag gccgttgggg ttagatttct ggattggctt accagcagag 600
 caagaggcac gggtcgcgcc gatgatcgcg gcggagcctg atccgcaaag cctcttcttc 660
 caagaggtcg cgaagcctgg ggccttacag tcgctcgtag tccttaactc cggcggctat 720
 atgggtgctc agcctgagta tgactcgcg gcggcgcatg cggccgagat tgggtgcagcc 780
 ggtggtatca ccaacgcacg cggcctggca ggcattgtac caccactggc ctgcggaggc 840
 aaactcaaag ggggtggagtt ggtcagtcct gacatgctgg cccgaatgtc cagagtggcc 900
 tctgcgactg ggagagatgc cgtgctcatg atgccaaccc ggtttgcctt gggcttcatg 960
 aagtccatgg acaaccgccg ggagcctgct ggcgtgcagg acagcgcgct ctttggggag 1020
 gaggcttttg gccatgtggg ggccgggggt tcgtttgggt ttgccgatcc caaagcagga 1080
 atgtcctttg gctataccat gaaccgaatg gggctgggag cggggctcaa cccgcggggg 1140
 caaagcctgg tggatgcaac ctaccgctcg ttagggatat agtcggatgc ctctggagcc 1200
 tggacctga 1209

<210> 4
 <211> 402
 <212> PRT
 <213> Unknown

<220>
 <223> Environmental

<220>
 <221> DOMAIN
 <222> (18)...(386)
 <223> Beta-lactamase

<400> 4
 Met Glu Ile His Gly Thr Cys Asp Pro Lys Phe His Leu Val Arg Gln
 1 5 10 15

Glu Phe Glu Arg Asn Leu Arg Glu Arg Gly Glu Val Gly Ala Ser Val
 20 25 30

Cys Val Thr Leu His Gly Glu Thr Val Val Asp Leu Trp Gly Gly Met
 35 40 45

Ala Arg Ala Asp Thr Gln Thr Pro Trp Thr Ala Glu Thr Val Ser Ile
 50 55 60

Val Phe Ser Ser Thr Lys Gly Ala Thr Ala Leu Cys Ala His Met Leu
 65 70 75 80

Ala Ser Arg Gly Gln Leu Asp Leu Asp Ala Pro Val Ala Thr Tyr Trp
 85 90 95

Pro Glu Phe Ala Gln Ala Gly Lys Ala Arg Ile Pro Val Lys Met Leu
 100 105 110

Leu Asn His Gln Ala Gly Leu Pro Ala Val Arg Thr Pro Leu Pro Gln
 115 120 125

Gly Ala Tyr Ala Asp Trp Glu Leu Met Val Asn Thr Leu Ala Lys Glu
 130 135 140

Glu Pro Phe Trp Glu Pro Gly Thr Arg Asn Gly Tyr His Ala Leu Thr
 145 150 155 160

Met Gly Trp Leu Val Gly Glu Val Val Arg Arg Val Ser Gly Lys Ser
 165 170 175

Leu Gly Thr Phe Phe Gln Glu Glu Ile Ala Arg Pro Leu Gly Leu Asp
 180 185 190

Phe Trp Ile Gly Leu Pro Ala Glu Gln Glu Ala Arg Val Ala Pro Met
 195 200 205

Ile Ala Ala Glu Pro Asp Pro Gln Ser Leu Phe Phe Gln Glu Val Ala
 210 215 220

Lys Pro Gly Ala Leu Gln Ser Leu Val Leu Leu Asn Ser Gly Gly Tyr
 225 230 235 240

Met Gly Ala Gln Pro Glu Tyr Asp Ser Arg Ala Ala His Ala Ala Glu
 245 250 255

Ile Gly Ala Ala Gly Gly Ile Thr Asn Ala Arg Gly Leu Ala Gly Met
 260 265 270

Tyr Ala Pro Leu Ala Cys Gly Gly Lys Leu Lys Gly Val Glu Leu Val
 275 280 285

Ser Pro Asp Met Leu Ala Arg Met Ser Arg Val Ala Ser Ala Thr Gly
 290 295 300

Arg Asp Ala Val Leu Met Met Pro Thr Arg Phe Ala Leu Gly Phe Met
 305 310 315 320

Lys Ser Met Asp Asn Arg Arg Glu Pro Ala Gly Val Gln Asp Ser Ala
 325 330 335

Leu Phe Gly Glu Glu Ala Phe Gly His Val Gly Ala Gly Gly Ser Phe
 340 345 350

Gly Phe Ala Asp Pro Lys Ala Gly Met Ser Phe Gly Tyr Thr Met Asn
 355 360 365

Arg Met Gly Leu Gly Ala Gly Leu Asn Pro Arg Gly Gln Ser Leu Val
 370 375 380

Asp Ala Thr Tyr Arg Ser Leu Gly Tyr Gln Ser Asp Ala Ser Gly Ala
 385 390 395 400

Trp Thr

<210> 5
 <211> 1578
 <212> DNA
 <213> Unknown

<220>
 <223> Environmental

<400> 5
 atgagatcag cagctcgcat cagcgtggcg gcagttgcct ttctttgcct gctcttgacg 60
 actcggggtt cgcgccagat cgtgccggcg atggaatgtg cggatctggc gaatcagcag 120
 cttcccaaca cgacgatcac ctgcggccag accgtcacca ccggatcggt aacgcccccg 180
 ggctcgacga atccgatcac cgacctgcct cctttctgcc gtgtcacagg cgccatcgcc 240
 ccgacgagcg agtcgcacat cctcttcgag gtctggctgc cgctggataa atggaacggc 300
 aagttcgccg gcgtgggcaa cggcggctgg gccggcatca tctccttcgg cgccctcgga 360
 agccagctca agcgcggcta cgcgaccgcc tccacgaata cgggtcacga agcggcgccg 420
 gggatgaacg cagccagggt tgcgttcgag aagccggagc agcttatcga cttcgcctat 480
 cgctcccagc acgagacggc cctgaaagcg aaggcgctgg ttcaggcttt ctacgggaag 540
 ccgccggaac actcctattt catcgggtgc tcatcggtg ggtaccaggg cctgatggag 600
 gcccaacgat ttccggccga ctacgacggg atcgtcgccg gtatgccggc gaacaactgg 660
 acacggctga tggccggcga cttggacgcg atccttgccg tctccgtaga tcctgccagc 720
 caccttcccg tctccgcatt gggctctgtt tatcgctcgg tgcctcgtgc ctgcgacggc 780
 atcgacggtg ttgtagacgg tgttctggag gatccgcgcc gatgccggtt cgacccggcc 840
 gtgttgatgt gcaaggcgga tcagaatccc gatggctgcc ttacgccggc tcagggtgaa 900
 gcggcacggc gcatatacgg cggctctgaag gatcccaaga ccggcgctca gctctatccg 960
 gggctggcgc cgggaagcga gccgttcttg ccgcaccgca atccggcgaa tccgttcctt 1020
 attccgatcg cgcactacaa gtggctcgtc ttgccgatc caaactggga ttggagaaca 1080
 ttcaagttca cggatccggc ggactaccag gctttcctca atgcggaagc cacgtatgcc 1140
 cctactctca atgcgaccaa tccggacctc cgggagttca gccggcgcgg cggcagggtg 1200
 attcagtacc atggctggaa cgatcagctg attgccccgc aaaacagcat cgactattac 1260
 gagagcgtcc ttctgttctt cgggtccggc aaacaggatc gagcgcagac cgtgcgcgag 1320
 gttcagagct tctaccggct gttcatggcg ccgggtatgg ctactgtgg aggcgggtaca 1380
 ggtccgaact catttgacat gctggatgcc ctcgagaagt ggggtggaagg cgggatagcg 1440
 ccggaacgag tccttgcgac gcgttcata aacggcgtag tcgaccggct gcgcccgtc 1500

tgtccatatc cgcaggtcgc cgtgtacaag ggtcatgggg atacaaacga cgccgcgaac 1560
 ttcgtctgtc gcgattag 1578

<210> 6
 <211> 525
 <212> PRT
 <213> Unkown

<220>
 <223> Environmental

<220>
 <221> SIGNAL
 <222> (1)...(25)

<400> 6
 Met Arg Ser Ala Ala Arg Ile Ser Val Ala Ala Val Ala Phe Leu Cys
 1 5 10 15
 Leu Leu Leu Thr Thr Arg Val Ser Ala Gln Ile Val Pro Ala Met Glu
 20 25 30
 Cys Ala Asp Leu Ala Asn Gln Gln Leu Pro Asn Thr Thr Ile Thr Ser
 35 40 45
 Ala Gln Thr Val Thr Thr Gly Ser Leu Thr Pro Pro Gly Ser Thr Asn
 50 55 60
 Pro Ile Thr Asp Leu Pro Pro Phe Cys Arg Val Thr Gly Ala Ile Ala
 65 70 75 80
 Pro Thr Ser Glu Ser His Ile Leu Phe Glu Val Trp Leu Pro Leu Asp
 85 90 95
 Lys Trp Asn Gly Lys Phe Ala Gly Val Gly Asn Gly Gly Trp Ala Gly
 100 105 110
 Ile Ile Ser Phe Gly Ala Leu Gly Ser Gln Leu Lys Arg Gly Tyr Ala
 115 120 125
 Thr Ala Ser Thr Asn Thr Gly His Glu Ala Ala Pro Gly Met Asn Ala
 130 135 140
 Ala Arg Phe Ala Phe Glu Lys Pro Glu Gln Leu Ile Asp Phe Ala Tyr
 145 150 155 160

Arg Ser Gln His Glu Thr Ala Leu Lys Ala Lys Ala Leu Val Gln Ala
 165 170 175

Phe Tyr Gly Lys Pro Pro Glu His Ser Tyr Phe Ile Gly Cys Ser Ser
 180 185 190

Gly Gly Tyr Gln Gly Leu Met Glu Ala Gln Arg Phe Pro Ala Asp Tyr
 195 200 205

Asp Gly Ile Val Ala Gly Met Pro Ala Asn Asn Trp Thr Arg Leu Met
 210 215 220

Ala Gly Asp Leu Asp Ala Ile Leu Ala Val Ser Val Asp Pro Ala Ser
 225 230 235 240

His Leu Pro Val Ser Ala Leu Gly Leu Leu Tyr Arg Ser Val Leu Ala
 245 250 255

Ala Cys Asp Gly Ile Asp Gly Val Val Asp Gly Val Leu Glu Asp Pro
 260 265 270

Arg Arg Cys Arg Phe Asp Pro Ala Val Leu Met Cys Lys Ala Asp Gln
 275 280 285

Asn Pro Asp Gly Cys Leu Thr Pro Ala Gln Val Glu Ala Ala Arg Arg
 290 295 300

Ile Tyr Gly Gly Leu Lys Asp Pro Lys Thr Gly Ala Gln Leu Tyr Pro
 305 310 315 320

Gly Leu Ala Pro Gly Ser Glu Pro Phe Trp Pro His Arg Asn Pro Ala
 325 330 335

Asn Pro Phe Pro Ile Pro Ile Ala His Tyr Lys Trp Leu Val Phe Ala
 340 345 350

Asp Pro Asn Trp Asp Trp Arg Thr Phe Lys Phe Thr Asp Pro Ala Asp
 355 360 365

Tyr Gln Ala Phe Leu Asn Ala Glu Ala Thr Tyr Ala Pro Thr Leu Asn
 370 375 380

Ala Thr Asn Pro Asp Leu Arg Glu Phe Ser Arg Arg Gly Gly Arg Leu
 385 390 395 400

Ile Gln Tyr His Gly Trp Asn Asp Gln Leu Ile Ala Pro Gln Asn Ser
405 410 415

Ile Asp Tyr Tyr Glu Ser Val Leu Ser Phe Phe Gly Ser Gly Lys Gln
420 425 430

Asp Arg Ala Gln Thr Val Arg Glu Val Gln Ser Phe Tyr Arg Leu Phe
435 440 445

Met Ala Pro Gly Met Ala His Cys Gly Gly Gly Thr Gly Pro Asn Ser
450 455 460

Phe Asp Met Leu Asp Ala Leu Glu Lys Trp Val Glu Gly Gly Ile Ala
465 470 475 480

Pro Glu Arg Val Leu Ala Thr Arg Ser Ile Asn Gly Val Val Asp Arg
485 490 495

Leu Arg Pro Leu Cys Pro Tyr Pro Gln Val Ala Val Tyr Lys Gly His
500 505 510

Gly Asp Thr Asn Asp Ala Ala Asn Phe Val Cys Arg Asp
515 520 525